SQL Saturday Boston 2023

Using Apache Kafka to replicate data from Postgres to SQL Server

achieving an heterogenous database replication solution

Marcelo Adade

Principal Consultant - Pythian

SQL Saturday (#1046)

Thank You! Sponsors

Connect, share, & learn with peers and thought leaders while celebrating all things data for a week of learning and networking opportunities.

(use **REDGATE4SUMMIT** to get U$ 150 off)

November 13th Through November 17th in Seattle, WA

https://passdatacommunitysummit.com/

SQL Saturday (#1046)

Sp_who(dba) – Marcelo Adade

- **Microsoft Data Platform Internal Principal Consultant at Pythian (Ottawa)**
  We're hiring!
- **Database Consultant and Instructor at DBBITS (Sao Paulo)**
- **Microsoft Data Platform MVP**
- **SQLSATURDAY organizer**
- https://linktr.ee/marceloadade
- https://pythian.com/careers/

What we'll talk about today

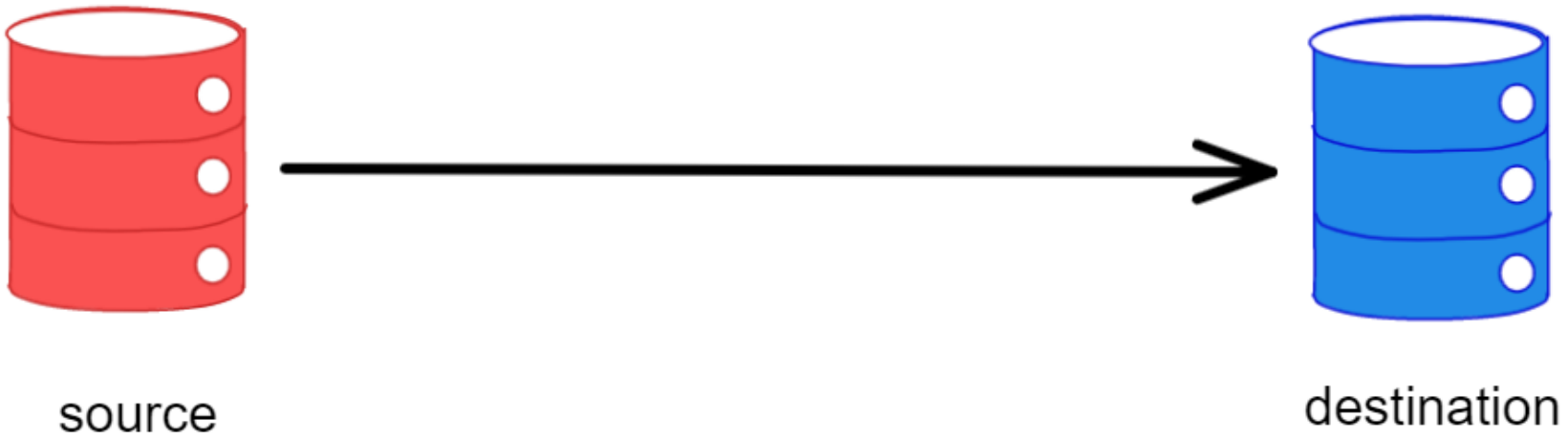
- **The replication challenge**
- **Possible solutions**
- **Confluent Kafka**

You'll learn

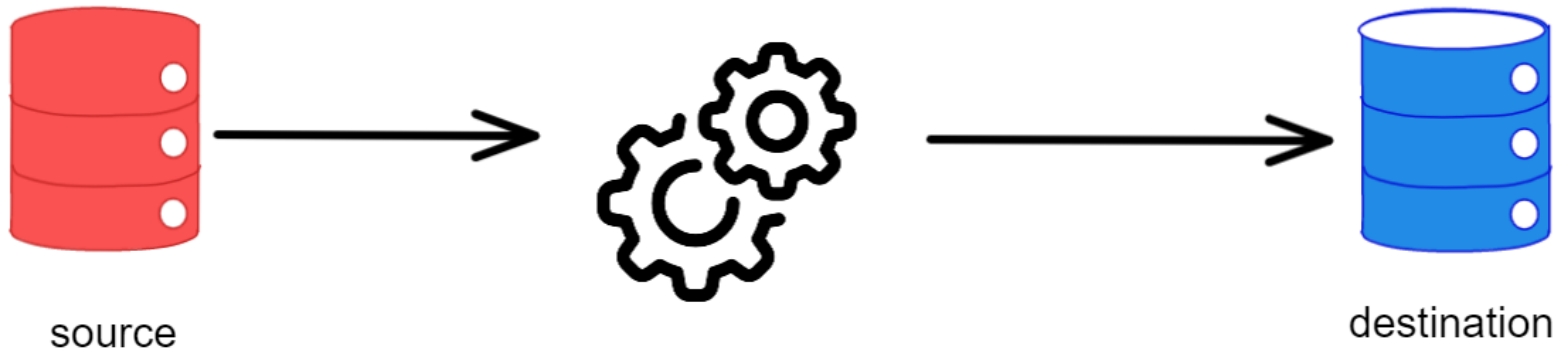
A possible (low cost) solution for achieving heterogeneous replication in your environment

The scenario

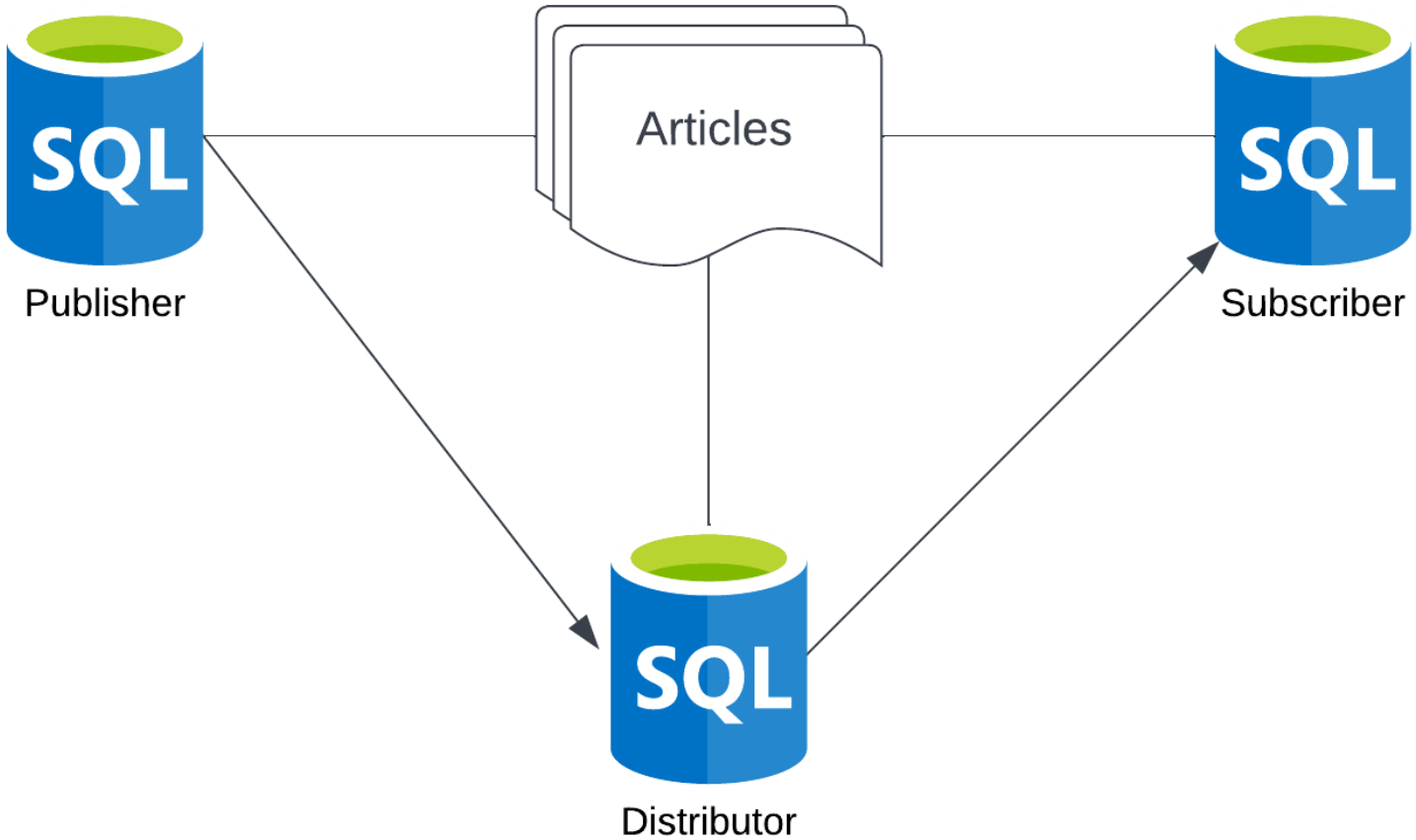
Copy data



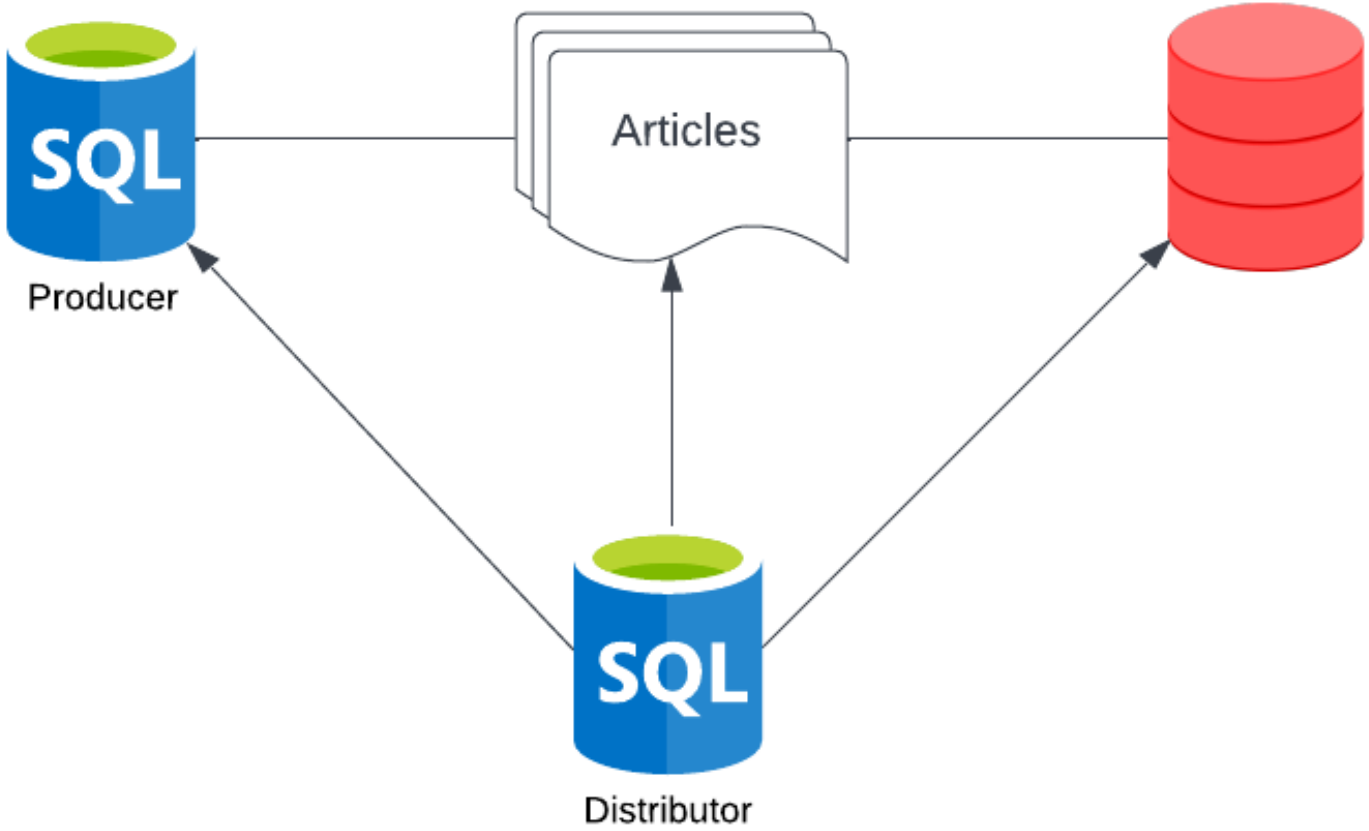
Copy data



Native Replication



Heterogeneous replication



The problem

- No official support
- Only Oracle was available
- No heterogeneous replication from the Pgsql side
- A third-party tool would be somewhat expensive

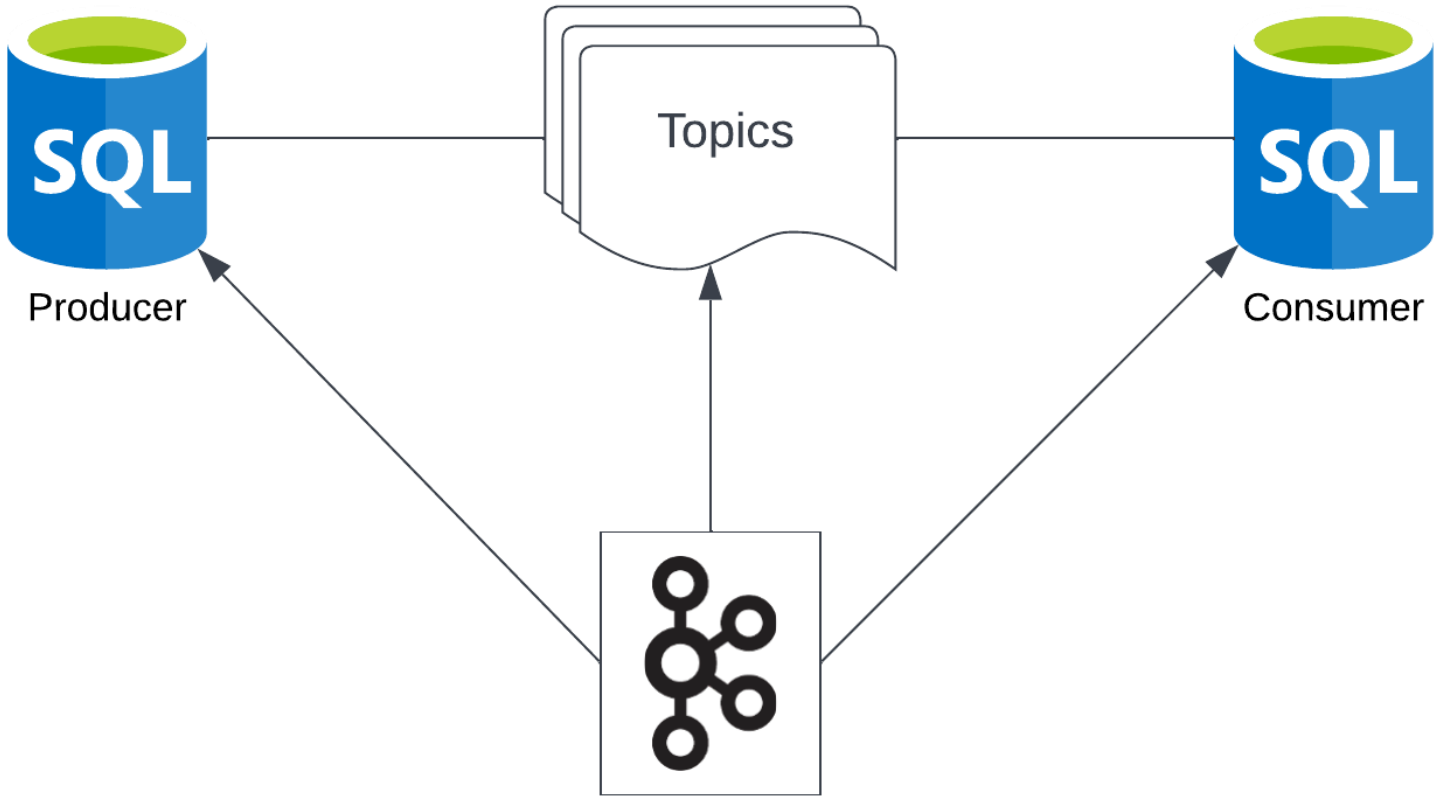
The (not so good to maintain) solution

- Triggers
- Procedures

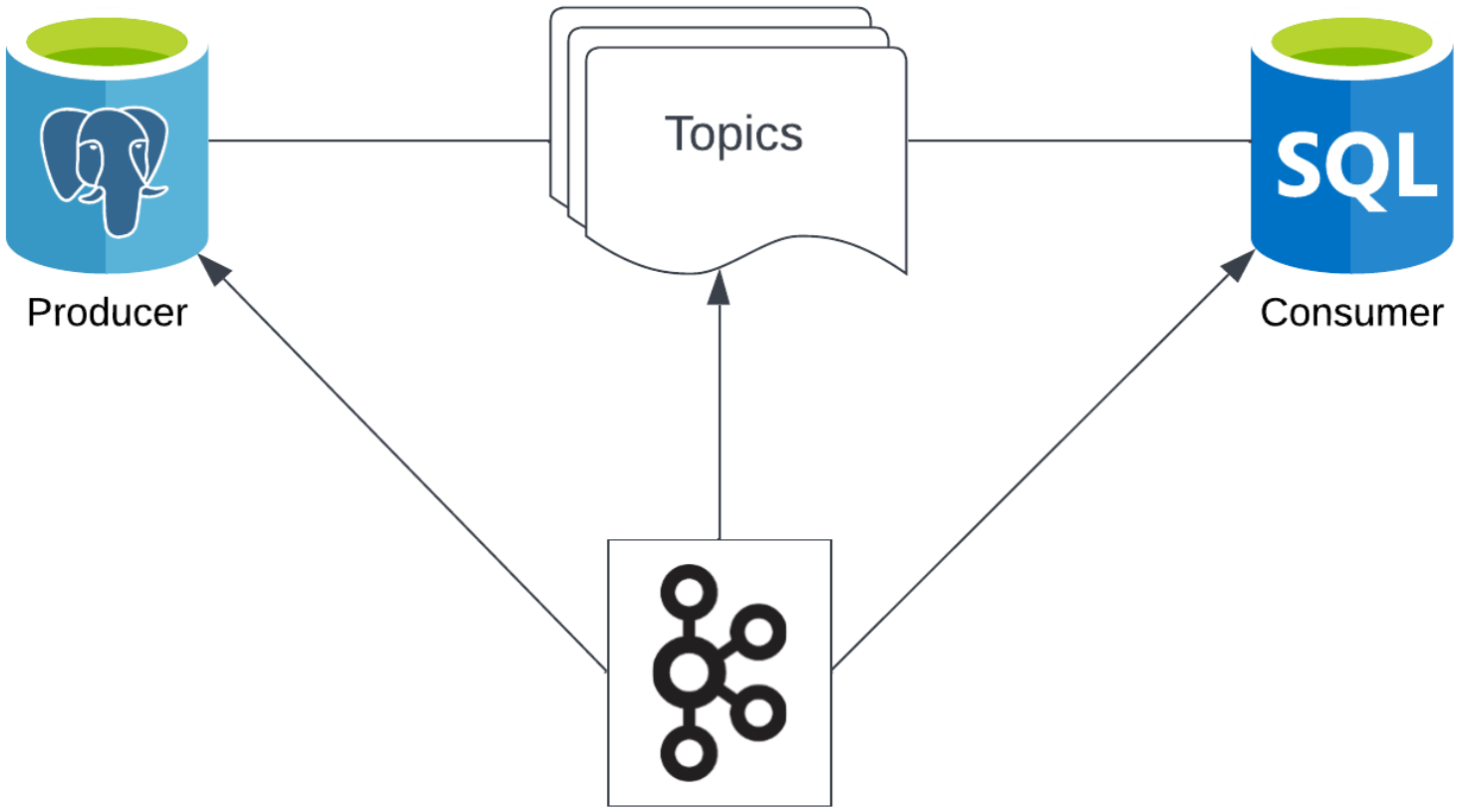
Possible $olutions

Another possible solution

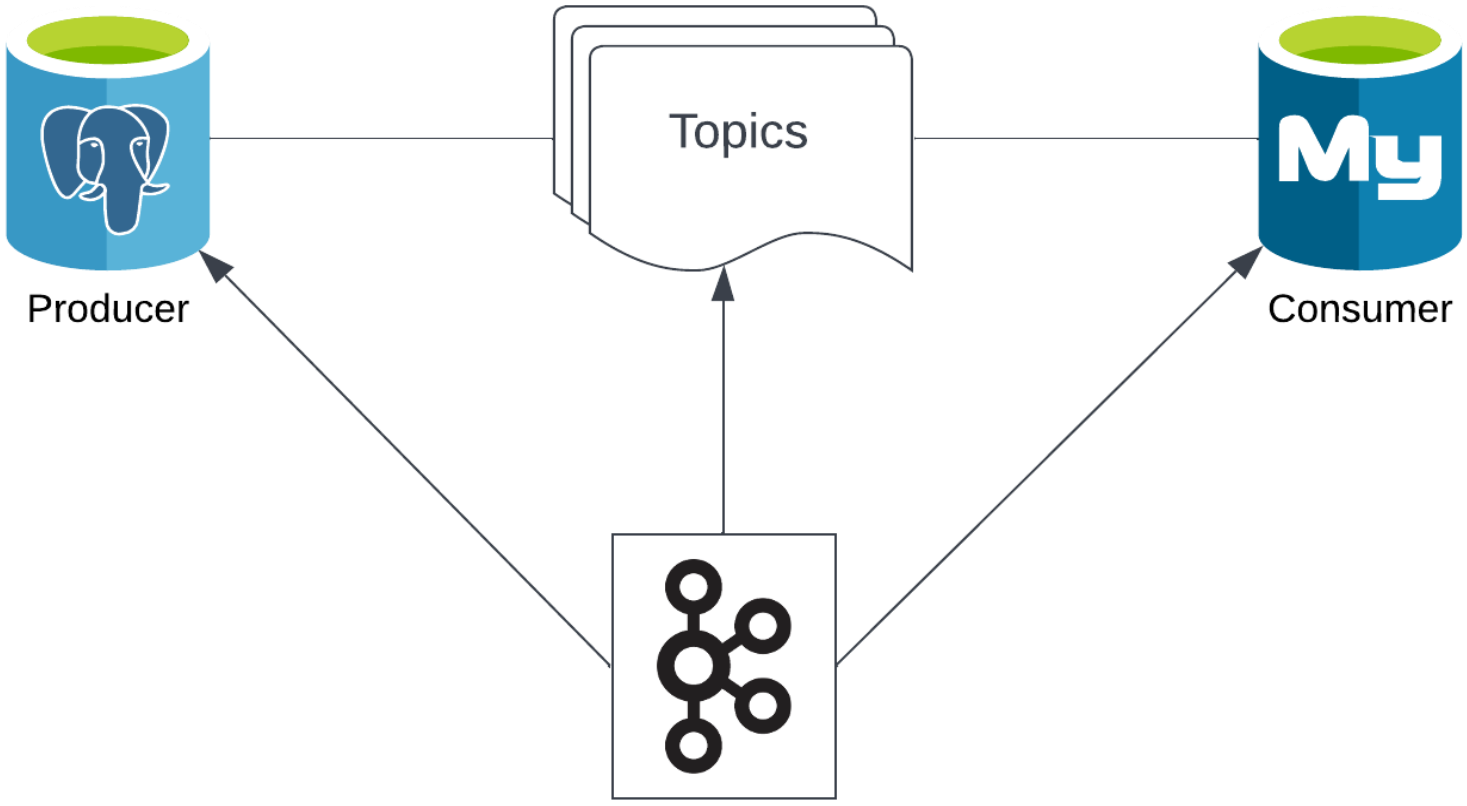
Enter Apache Kafka



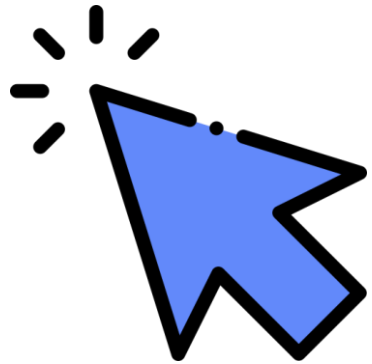
Enter Apache Kafka



Enter Apache Kafka



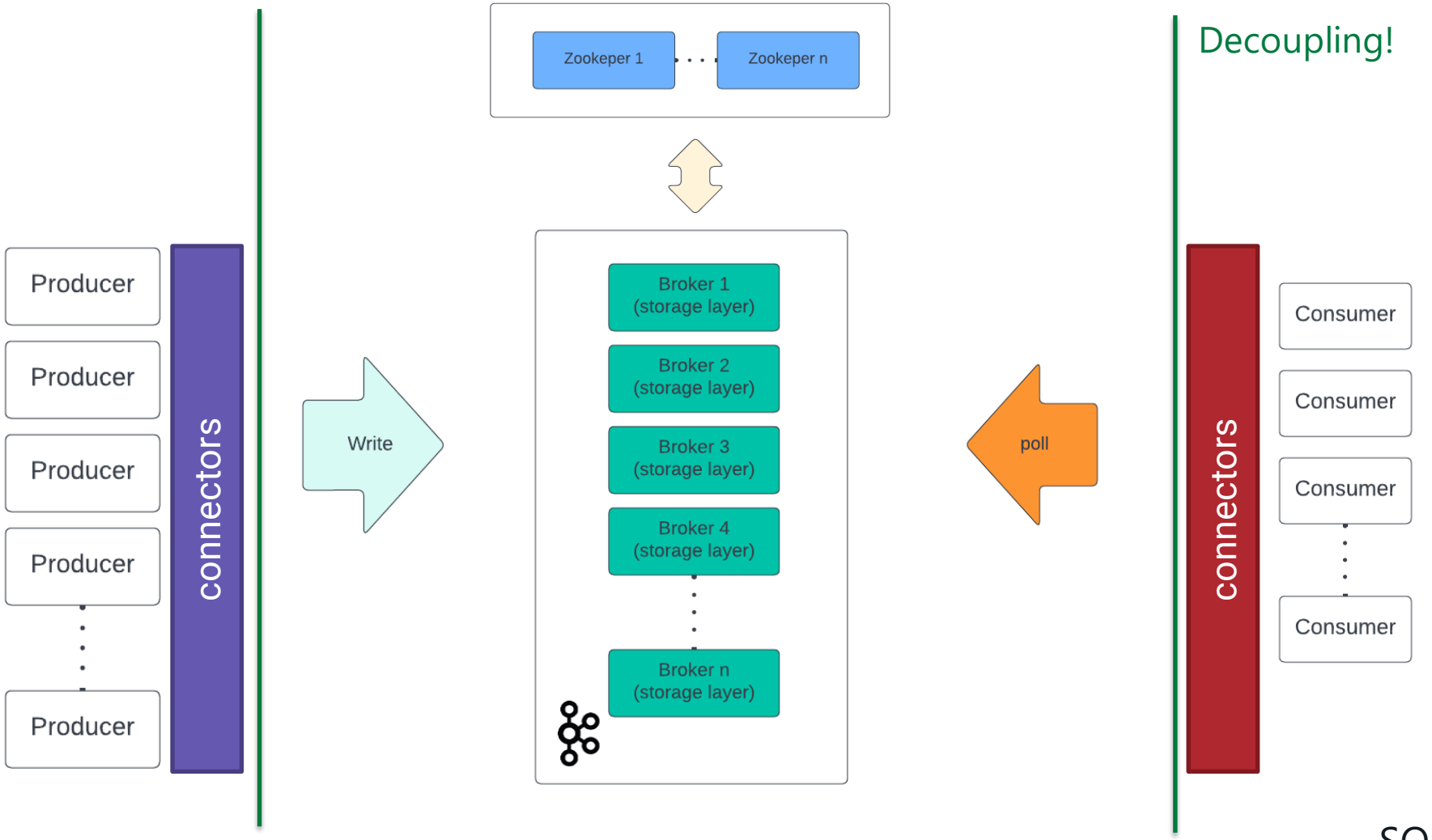
Apache Kafka is great for:



Who uses it

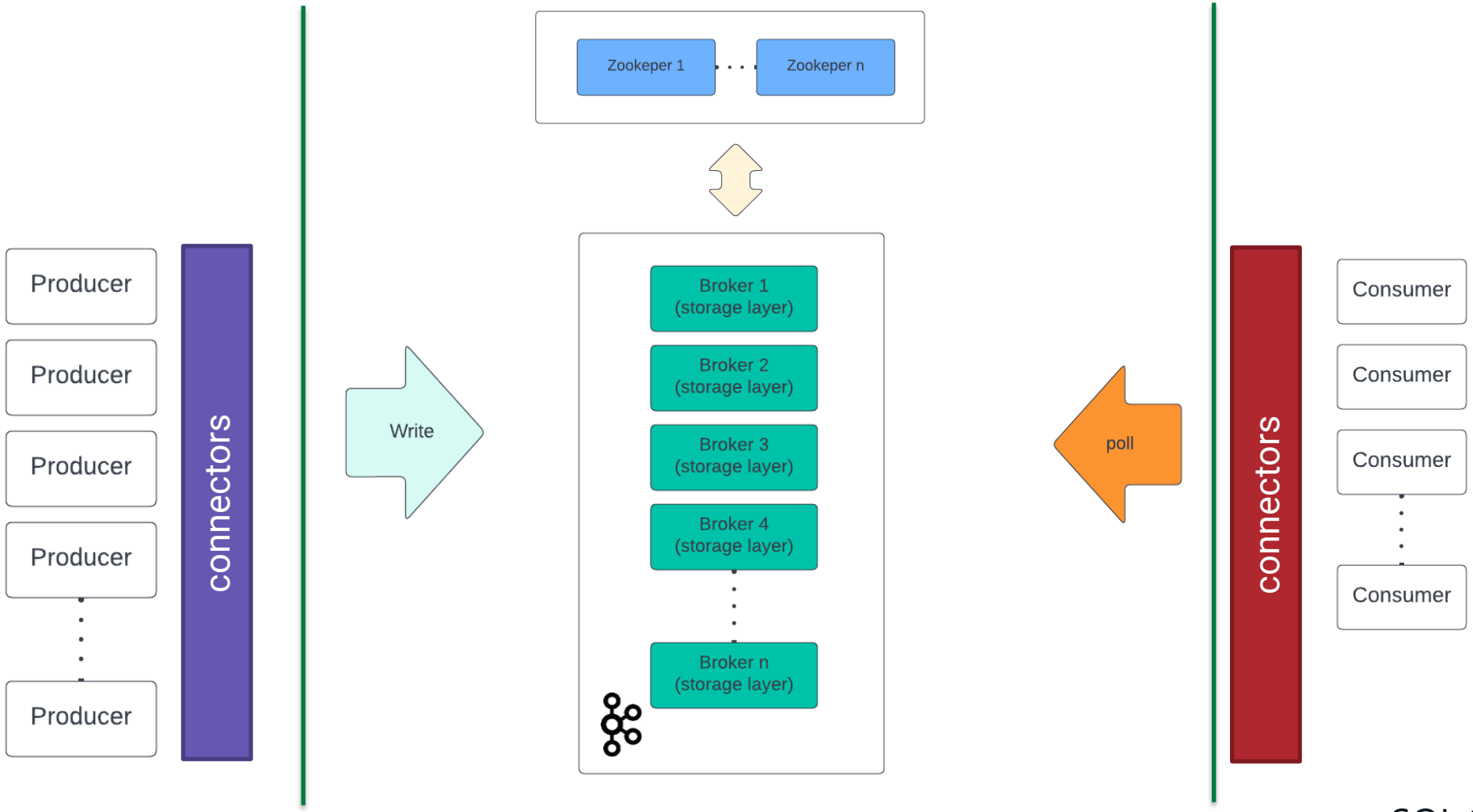
https://kafka.apache.org/powered-by

Apache Kafka components



Decoupling!

Database special case

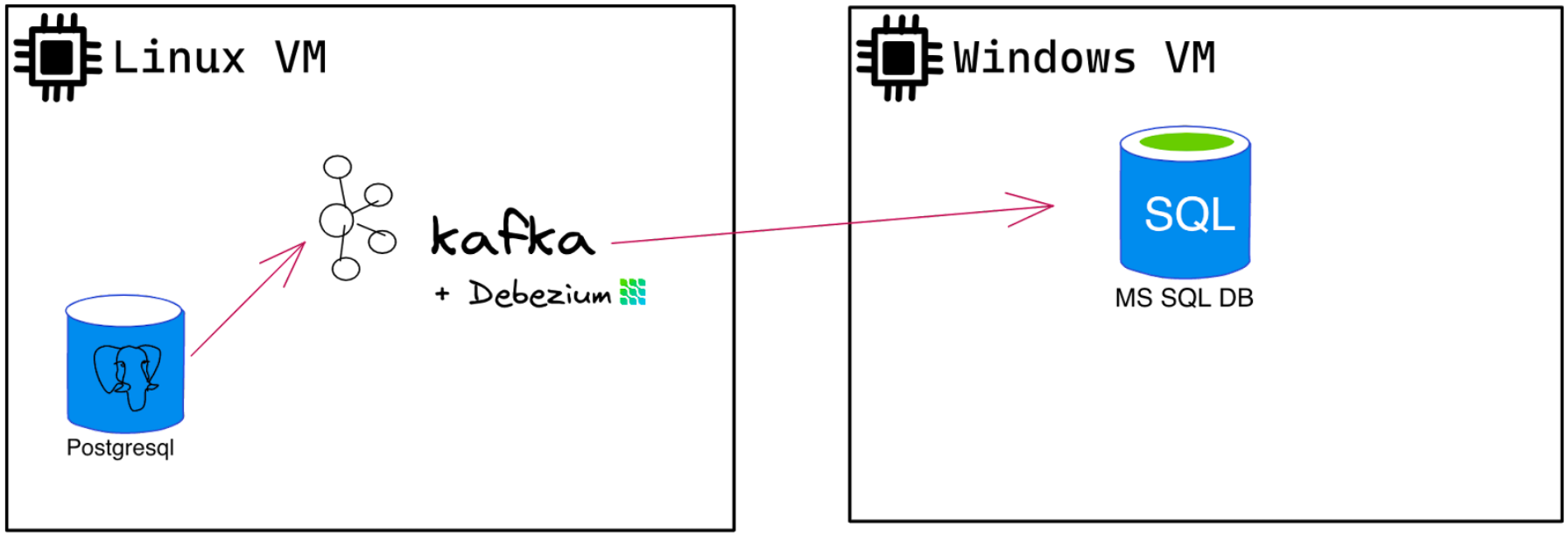


Solution 1 (“Manual” Kafka)

Manual Step-by-step

1. Define a data source or Producer
2. Define a data destination or Consumer
3. Install the Confluent Kafka platform
4. Install the required drivers
5. Create the connectors using the drivers
6. Monitor the Data flow
7. Sit back and relax!

Demo on manual Kafka



Solution 2 (Cloud Kafka)

Confluent Cloud

CONFLUENT

Google Cloud

https://www.confluent.io/confluent-cloud/

Recap

3 possible solutions for replicating Data thru different Database technologies:

- Manually pushing data via triggers and procs
- Using Apache Kafka and setting it up manually
- Using Confluent Cloud within your provider of choice

References

- https://kirti-garg0410.medium.com/apache-kafka-architecture-kafka-component-overview-ac7acfd0ec85
- https://kafka.apache.org/intro
- https://learn.microsoft.com/en-us/sql/relational-databases/replication/sql-server-replication?view=sql-server-ver16
- https://debezium.io/documentation/reference/stable/connectors/postgresql.html#setting-up-postgresql
- https://docs.confluent.io/kafka-connect-jdbc/current/index.html
- https://learn.microsoft.com/en-us/sql/relational-databases/replication/non-sql/heterogeneous-database-replication?view=sql-server-ver16

Thank you and keep in touch!

- https://linktr.ee/marceloadade
- https://pythian.com/careers/

Pythian
love your data®